

Specific Motifs Recognized by the SH2 Domains of Csk, 3BP2, fps/fes, GRB-2, HCP, SHC, Syk, and Vav

Z. SONGYANG,^{1,2,3} S. E. SHOELSON,^{4,5} J. McGLADE,⁶ P. OLIVIER,⁶ T. PAWSON,⁶ X. R. BUSTELO,⁷
M. BARBACID,⁷ H. SABE,⁸ H. HANAFUSA,⁸ T. YI,⁹ R. REN,⁸ D. BALTIMORE,⁸
S. RATNOFSKY,¹⁰ R. A. FELDMAN,^{11,12} AND L. C. CANTLEY^{1,3*}

Department of Cell Biology, Harvard Medical School,¹ Department of Medicine, Beth Israel Hospital,² Research Division, Joslin Diabetes Center,³ and Department of Medicine, Brigham and Women's Hospital and Harvard Medical School,⁵ Boston, Massachusetts 02115; Division of Molecular and Developmental Biology, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada⁶; Department of Molecular Biology and Molecular Sciences, Bristol-Myers Squibb Pharmaceutical Research Institute, Princeton, New Jersey 08543⁷; Rockefeller University, New York, New York 10021⁸; Department of Cancer Biology, Research Institute, The Cleveland Clinic Foundation, Cleveland, Ohio 44195⁹; BASF BioResearch Corporation, Worcester, Massachusetts 01605¹⁰; Department of Microbiology and Immunology, University of Maryland School of Medicine,¹¹ and Medical Biotechnology Center, University of Maryland Biotechnology Institute,¹² Baltimore, Maryland 21201; and Department of Physiology, Tufts University School of Medicine, Boston, Massachusetts 02111³

Received 27 October 1993/Returned for modification 3 December 1993/Accepted 10 January 1994

Src homology 2 (SH2) domains provide specificity to intracellular signaling by binding to specific phosphotyrosine (phospho-Tyr)-containing sequences. We recently developed a technique using a degenerate phosphopeptide library to predict the specificity of individual SH2 domains (*src* family members, Abl, Nck, Sem5, phospholipase C- γ , p85 subunit of phosphatidylinositol-3-kinase, and SHPTP2 (Z. Songyang, S. E. Shoelson, M. Chaudhuri, G. Gish, T. Pawson, W. G. Haser, F. King, T. Roberts, S. Ratnoffsky, R. J. Lechleider, B. G. Neel, R. B. Birge, J. E. Fajardo, M. M. Chou, H. Hanafusa, B. Schaffhausen, and L. C. Cantley, Cell, 72:767-778, 1993). We report here the optimal recognition motifs for SH2 domains from GRB-2, Drk, Csk, Vav, fps/fes, SHC, Syk (carboxy-terminal SH2), 3BP2, and HCP (amino-terminal SH2 domain, also called PTP1C and SHPTP1). As predicted, SH2 domains from proteins that fall into group I on the basis of a Phe or Tyr at the β D5 position (GRB-2, 3BP2, Csk, fps/fes, Syk C-terminal SH2) select phosphopeptides with the general motif phospho-Tyr-hydrophilic (residue)-hydrophilic (residue)-hydrophobic (residue). The SH2 domains of SHC and HCP (group III proteins with Ile, Leu, or Cys at the β D5 position) selected the general motif phospho-Tyr-hydrophobic-Xxx-hydrophobic, also as predicted. Vav, which has a Thr at the β D5 position, selected phospho-Tyr-Met-Glu-Pro as the optimal motif. Each SH2 domain selected a unique optimal motif distinct from motifs previously determined for other SH2 domains. These motifs are used to predict potential sites in signaling proteins for interaction with specific SH2 domain-containing proteins. The Syk SH2 domain is predicted to bind to Tyr-hydrophilic-hydrophilic-Leu/Ile motifs like those repeated at 10-residue intervals in T- and B-cell receptor-associated proteins. SHC is predicted to bind to a subgroup of these same motifs. A structural basis for the association of Csk with Src family members is also suggested from these studies.

The activation of cellular protein tyrosine kinases by growth factors, lymphokines, and cytokines initiates a cascade of events critical for mitosis and other cellular responses. Transphosphorylation or autophosphorylation of receptors on tyrosine creates binding sites for cytoplasmic proteins that are involved in signal transduction (4). Cytosolic proteins known to be involved in protein tyrosine kinase signaling cascades include phosphatidylinositol 3-kinase (PI-3-kinase) (17, 50), phospholipase C- γ (PLC- γ) (27), SHC (33), GRB-2 (32), Ras-GTP-activating protein (16), and pp60^{c-src} and *src*-like protein tyrosine kinases (21). These molecules all contain regions of homology to a noncatalytic domain in pp60^{c-src} called the Src homology 2 (SH2) domain. SH2 domains alone are sufficient to confer direct binding to various activated receptors, and this binding depends on tyrosine phosphorylation of the receptors (1, 4). On the basis of a short sequence of

similarity between sites of tyrosine phosphorylation on polyomavirus middle T antigen (51) and the platelet-derived growth factor (PDGF) receptor (10, 19) that are involved in binding PI-3-kinase, it was proposed that individual SH2 domains recognize phosphotyrosine (phospho-Tyr) residues within distinct primary sequences (4). This concept has been supported because separate sites for binding of SH2 domains from different signaling molecules have been mapped on various receptors (11, 23, 26, 36, 40, 41, 48).

Recently, we developed a procedure for rapidly predicting the optimal phosphopeptide motif for any SH2 domain (45). From a comparison of consensus sequences of *in vivo* binding sites of various SH2 domain-containing proteins, it appeared that residues immediately carboxy terminal of the phospho-Tyr were likely to regulate specificity. We designed a phosphopeptide library with the sequence Gly-Asp-Gly-phospho-Tyr-Xxx-Xxx-Xxx-Ser-Pro-Leu-Leu-Leu, where Xxx indicates positions of degeneracy in which all amino acids except Trp or Cys are possible. Immobilized forms of SH2 domains from various proteins were used to affinity purify the peptides preferred by each SH2 domain. By sequencing the mixture of peptides rather than purifying individual peptides, it was

* Corresponding author. Mailing address: Warren Alpert 152, Harvard Medical School, 200 Longwood Ave., Boston, MA 02115. Phone: (617) 278-3040. Fax: (617) 278-3033.

† Present address: Department of Physiology, Tufts University School of Medicine, Boston, MA 02111.

possible to determine preferred amino acids at the positions +1, +2, and +3 carboxy terminal of the phospho-Tyr. By this technique, the optimal sequence for the p85 N-terminal SH2 (N-SH2) domain was predicted to be phospho-Tyr-Met/Val/Ile/Glu-Xxx-Met, in good agreement with the sites found to bind PI-3-kinase *in vivo*. The optimal motifs for 12 other SH2 domains were also predicted. It was noted that while SH2 domains from the Src family of protein tyrosine kinases selected identical optimal motifs (phospho-Tyr-Glu-Glu-Ile), all other SH2 domains selected unique optimal motifs. On the basis of these motifs, potential sites for binding of SH2-containing proteins in receptors and other signaling molecules were predicted. Several of these predictions have been confirmed (20, 22, 32, 44, 47).

While this research was in progress, Waksman et al. (49) determined the crystal structures of the Src SH2 domain associated with two different low-affinity phosphopeptides. These structures provided an explanation for the ability of SH2 domains to specifically bind phospho-Tyr rather than free Tyr or other phosphoamino acids. However, since the side chains of the associated peptides did not make significant contact with the SH2 domain, the basis for sequence-specific selection was not apparent. On the basis of our results showing that the Src and Lck SH2 domains have high affinity for the sequence phospho-Tyr-Glu-Glu-Ile, Waksman et al. (using Src SH2) and Eck et al. (using Lck SH2) attempted cocrystallizations with peptides containing this motif (9, 49). The crystal structures revealed an explanation for the selection of this motif. The Ile at the +3 position fit into a hydrophobic pocket that was induced during the binding. The Glu residues at the +1 and +2 positions lay on the surface of the SH2 domain and formed hydrophobic contacts and salt bonds with side chains of residues from the SH2 domain.

Consistent with our observation that members of the Src family of SH2 domains selected identical optimal motifs, the side chains predicted to make contact with the associated peptides were conserved in this family. In contrast, all other SH2 domains had distinct amino acid replacements at one or more of the contact residues. These substitutions offered an explanation for the distinct motifs selected by these other SH2 domains. We noted that the amino acid at the β D5 position of various SH2 domains is likely to have a major influence on selectivity, since in the Src and Lck structures, the side chain of this residue makes contact with the side chains of both the +1 Glu and the +3 Ile of the associated phosphopeptide (45). We placed the known SH2 domains into four groups on the basis of the amino acid at the β D5 position. The group 1 SH2 domains, which include the Src family, have Tyr or Phe at the β D5 position and select phosphopeptides with the general motif phospho-Tyr-hydrophilic (residue)-hydrophilic (residue)-hydrophobic (residue). Group 3 SH2 domains have Cys, Ile, or Leu at the β D5 position and select for the general motif phospho-Tyr-hydrophobic-Xxx-hydrophobic.

We report here the identification of sequence motifs recognized by the SH2 domains of GRB-2, Drk (31), 3BP2 (37), SHC (33), HCP (N terminal) (52), Syk (C terminal), Csk (43), fps/fes (12, 15), and Vav (18). Each domain selects a unique motif, although the general motifs of the proteins from groups I and III are consistent with the predictions by Songyang et al. (45). The Vav SH2 domain is thus far the sole member of the group II family (Thr at β D5), and its motif was determined to be phospho-Tyr-Met-Glu-Pro. The identification of these motifs provides additional evidence for the structural basis of specific recognition by SH2 domains. Potential cellular targets for the proteins bearing these motifs are predicted.

MATERIALS AND METHODS

SH2 GST fusion proteins. cDNAs encoding SH2 domains of mouse 3BP2, GRB-2 (amino acid residues 59 to 160), HCP N terminal (residues 1 to 110), chicken Csk (residues 9 to 189), DGRB-2/Drk (full length), human fps/fes (residues 425 to 609), SHC (residues 344 to 473), Syk C terminal (residues 171 to 261), and Vav (residues 655 to 774) were subcloned into PGEX vectors and expressed as glutathione *S*-transferase (GST) fusion proteins. The fusion proteins were then produced in *Escherichia coli* by IPTG (isopropyl- β -D-thiogalactopyranoside) induction and purified with glutathione-agarose beads.

Phosphopeptide library design and synthesis. The phosphopeptide library used contains peptides with the general sequence GDGpYXXXSPLLL (single-letter amino acid designation), in which pY is phospho-Tyr and X indicates a position of degeneracy in which all amino acids but Cys or Trp have been included. The synthesis and analysis of this library were performed as previously described (45).

Affinity purification of phosphopeptides from a library. One hundred fifty microliters of glutathione-agarose beads containing approximately 300 μ g of the GST fusion protein of interest was packed in a 1-ml syringe as an affinity column. The beads were washed with 1 ml of phosphate-buffered saline (PBS) (150 mM NaCl, 3 mM KCl, 10 mM Na₂HPO₄, 2 mM KH₂PO₄ [pH 7.2]). Approximately 0.3 mg of the degenerate peptide mixture was loaded on the top of the column, and the column was allowed to sit at room temperature for 10 min. During this period, the column flow was stopped. The column was then quickly washed twice with 1 ml of ice-cold PBS (containing 10 mg of blue dextran per ml and 0.5% Nonidet P-40) and was washed once with 1 ml of ice-cold PBS (without blue dextran or detergent), with a plunger used to force the solution through. To elute the peptides, 200 μ l of 20 mM sodium phenylphosphate solution (pH 7.8) was added to the column at room temperature. All of the flowthrough was collected, dried to a small volume, and sequenced on the Applied Biosystems 477A Protein Sequencer.

Data analysis. The preference of amino acid residues at the degenerate positions near the phosphorylated tyrosine was determined by dividing the amount of each individual amino acid at a given sequencing cycle in the phenylphosphate eluant by the amount of the same amino acid in a control experiment (in which GST alone was used in the column purification step described above). This corrects for the small variation in relative abundance of the 18 amino acids at each degenerate position and for differences in yield of different amino acids during sequencing (45). In order to compare relative selectivities at different positions in a single sequence (Fig. 1), the data were normalized so that the sum of the quantities of amino acid abundance in a given cycle equals 18 (the number of distinct amino acids in the mixture). This means that if the relative abundance of amino acids detected at a particular cycle is the same in the affinity-purified peptide mixture as in the control experiment, then all amino acids will have values of 1. However, these values don't represent the real enrichment values because the nonspecific binding to GST has not been subtracted. To calculate the enrichment values presented in Table 1, the background of nonspecifically bound peptides was subtracted prior to the normalization.

RESULTS

Identification of motifs recognized by GRB-2 and DGRB-2/Drk SH2 domains. In order to determine how well the motifs

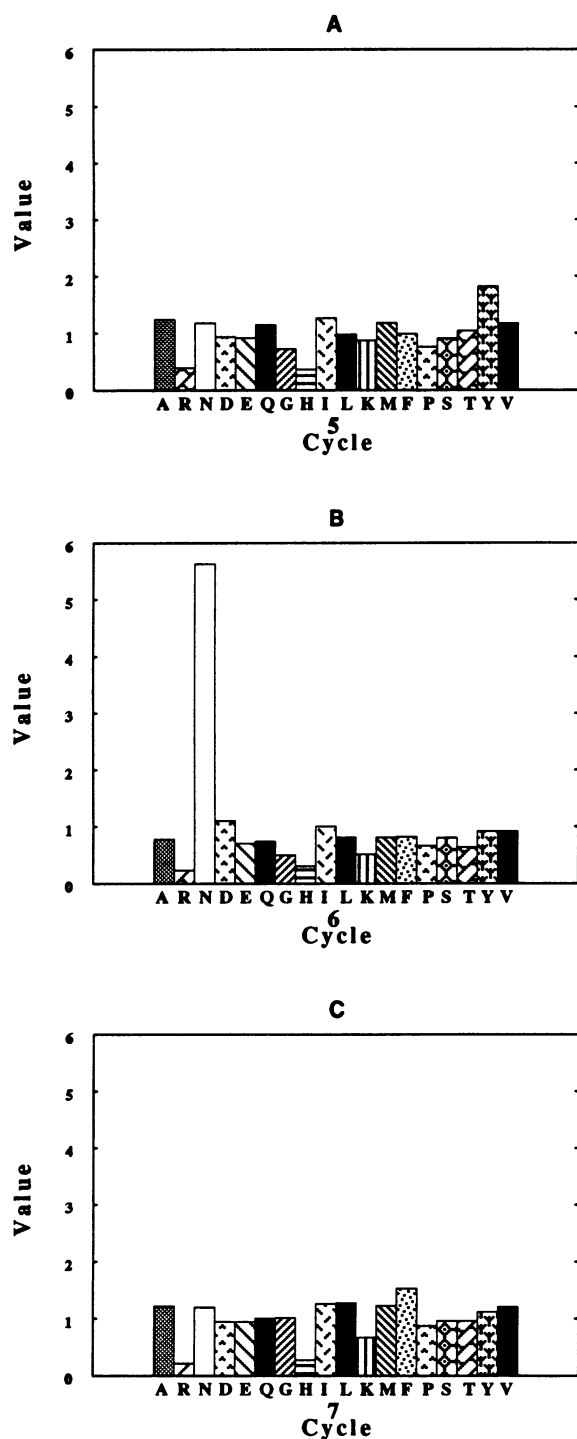


FIG. 1. Selection of phosphopeptides that bind the DGRB-2 domain. The C-terminal degenerate phosphopeptide mixture (45) was added to a column containing the SH2 domain of DGRB-2 as a GST fusion protein. The column was washed, and bound peptides were eluted with phenylphosphate. The eluted peptide mixture was subjected to microsequence analysis, and the results were compared with those from the eluate of a control column containing GST alone. Panels A, B, and C show results from the 5th, 6th, and 7th cycles of the sequence (i.e., the 1st, 2nd, and 3rd positions, respectively, after phospho-Tyr). Values (not the enrichment value [see Materials and Methods]) represent the ratio of the amount of each amino acid eluted from GST-SH2 bead columns divided by that of the control GST bead columns at the same cycle. Abbreviations for amino acid residues are

TABLE 1. Recognition specificities of SH2 domains^a

SH2 domain	pY+1	pY+2	pY+3	Selectivity
Csk	T (4.5) A (4.1) S (2.2)	K (1.9) R (1.7) Q (1.6) N (1.5)	M (2.7) I (2.5) V (2.3) R (1.6)	23
3BP2	E (8.0) M (2.2) V (1.7)	N (4.1) V (2.6) I (2.3)	X	33
fps/fes	E (3.6)	X	V (1.9) I (1.6)	7
DGRB-2	Y (2.3) I (1.4) V (1.3)	N (7.1)	F (1.7) L (1.3) I (1.3) V (1.3)	28
GRB-2	Q (3.0) Y (2.3) V (1.7)	N (5.8)	Y (2.2) Q (2.1) F (1.8)	39
HCP N terminal	F (2.5)	X	F (2.0) P (1.6) L (1.4) Y (1.4)	5
SHC	I (2.1) E (2.0) Y (1.7) L (1.6)	X	I (3.4) L (3.2) M (3.2)	7
Syk C terminal	Q (2.2) T (2.0) E (1.8)	E (1.7) Q (1.7)	L (3.4) I (1.7)	13
Vav	M (4.3) L (2.4) E (2.3)	E (2.3)	P (2.3)	23

^a pY + 1, pY + 2, and pY + 3 represent the amino acids preferentially selected by the indicated SH2 domains at positions +1, +2, and +3, respectively, carboxy terminal to the phospho-Tyr residue of the associated phosphopeptide mixture. The numbers in parentheses indicate the enrichment value of the amino acids selected. The calculation of these numbers is described in Materials and Methods (values less than 1.3 are not shown). Those amino acids with enrichment values greater than 3 are written in boldface. Selectivity was calculated by multiplying the enrichment values together.

selected by SH2 domains from divergent species are conserved, we compared the optimal peptides selected by the mouse GRB-2 and *Drosophila melanogaster* GRB-2/Drk SH2 domains with those previously shown to be selected by the *Candida elegans* homolog Sem5. In all three cases, the selectivity was primarily determined by the presence of an Asn at the +2 position carboxy terminal of the phospho-Tyr. The data for *Drosophila* GRB2 are presented in Fig. 1. It is apparent that, in agreement with the results previously published with Sem5 (45), Asn at the +2 position is very strongly selected, even compared with the closely related amino acids Asp and Gln. The amino acids selected at the +1 and +3 positions were also similar to those selected at this position by Sem5 (hydrophobic

as follows: A, Ala; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; Y, Tyr.

residue with Val or aromatic amino acid preferred). However, GRB-2 differed from Sem5 in also selecting for Gln at these sites. The optimal motif is presented in Table 1 along with the selectivity (the predicted fold difference in the affinities of the optimal motif and the average-affinity motif in the library). The optimal motif selected by the DGRB-2/Drk SH2 domain is also presented in Table 1. Again, the optimal motif is Asn at +2 and a hydrophobic residue (Val or an aromatic amino acid preferred) at +1 and +3. Thus, in agreement with the fact that the residues predicted to interact with the side chains of the bound phosphopeptide are conserved on this protein across species (Table 2), the motif selected is conserved. The strong selection for Asn at +2 is common to several members of the group IB family of SH2 domains.

Motifs for other group IB SH2 domains. In addition to the GRB2 subfamily, four other SH2 domains from the group IB family were investigated: Csk, fps/fes, Syk C-terminal SH2 (C-SH2), and 3BP2. The results are summarized in Table 1. As predicted, they tended to select for the general motif phospho-Tyr-hydrophilic-hydrophilic-hydrophobic. Each had a unique optimal motif. 3BP2 was like GRB-2 and Abl in having strong selection for Asn at the +2 position. However, it differed from these other two in having very strong selection for Glu at +1 and little selection at the +3 position. 3BP2 and Abl (which has somewhat weaker selection for Glu at +1 and a selection for Pro at +3 [described below]) are likely to have overlapping targets *in vivo*.

The fps/fes SH2 domain also selected strongly on the basis of Glu at the +1 position. However, unlike most other members of group I, fps/fes showed little selection at the +2 position. Val was weakly selected at the +3 position.

The C-SH2 domain of Syk selected peptides with the motif phospho-Tyr-Gln/Thr/Glu-Glu/Gln-Leu. The subgroup of SH2 domains present in Syk and the related T-cell protein ZAP70 are the closest members to Src SH2 domains outside the Src family. On the basis of the similarity at the predicted binding sites and a substitution of Pro for Ser/Thr at the +3 pocket, we suggested that these SH2 domains may bind a motif similar to that of Src, but selecting for Leu rather than Ile at the +3 position (45). The results are in agreement with this prediction and suggest that the Reth motifs (YXXL-X₇-YXXL) in T- and B-cell receptor-associated proteins evolved for optimal association with the tandem SH2 domains found in ZAP70 and Syk (see Discussion).

The other SH2 domain from group I that was investigated was the Csk SH2 domain. This domain was unusual in having a strong selection for Thr at the +1 position (Table 1). In addition, Csk is the only SH2 domain thus far investigated that selects for Arg at the +3 position (although Met is slightly preferred). As discussed below, this motif suggests a mechanism by which Csk selects Src family members to phosphorylate.

Vav, a group II SH2 domain member. It was of interest to determine the motif selected by Vav, since this is the only SH2 domain with a Thr residue at the β D5 position (group II) known to date. The optimal motif (phospho-Tyr-Met-Glu-Pro) and selectivity (36) are presented in Table 1. The strongest selection is for Met at the +1 position.

Group III SH2 domains. Two group III SH2 domains were investigated. The N-SH2 domain of hematopoietic cell phosphatase (HCP), like that of SHPTP2/syp, showed broad selection for phospho-Tyr-hydrophobic-Xxx-hydrophobic. However, unlike the results for the N-SH2 domain of SHPTP2 in which Val and Ile were the preferred hydrophobic residues at the +1 and +3 positions (Table 2), Phe was slightly preferred at +1 and +3 (Table 1). On the basis of the residues of the

N-SH2 domains of HCP and SHPTP2 predicted to be at the binding site of the peptide (only one difference [Table 2]), similar motifs are expected.

The SHC SH2 domain has Leu rather than Cys or Ile at the β D5 position but, on the basis of selectivity, behaves like a group III member. The general motif is phospho-Tyr-hydrophobic/Glu-X-Ile/Leu/Met. The strongest selection is at the +3 position. The Ile/Leu selection at the +3 position and relatively broad selection at the +1 and +2 positions suggest that SHC (like Syk) could interact with the Reth motifs in T and B cells (see Discussion).

DISCUSSION

Structural basis for phosphopeptide selection by SH2 domains. From the results in this paper and previous results of Songyang et al. (45), we have determined the motifs for 22 SH2 domains. These results are summarized in Table 2, in which the various SH2 domains are grouped according to the amino acid at the position analogous to the β D5 position of the Src SH2 domain. In the Src and Lck SH2 domain crystal structures, the Tyr residue at β D5 makes contact with the side chains of the +1 Glu and +3 Ile (carboxy terminal of the phospho-Tyr moiety) of the associated phosphopeptide (9, 49). Thus, the amino acid at this position is expected to play a prominent role in determining what phosphopeptides are selected. Approximately half of the known SH2 domains have a Tyr or Phe residue at the β D5 position (group I). We noted previously that the members of this group that we had analyzed at that time selected for peptides with the general motif phospho-Tyr-hydrophilic-hydrophilic-hydrophobic (45). The six additional SH2 domains from this group that we analyzed for this paper also tended to select this general motif, although each had a unique preference.

Interestingly, several of the group IB SH2 domains, including Abl, the Sem5/GRB-2/Drk family, and 3BP2 strongly selected for phosphopeptides with Asn at the +2 position. The structural basis for this selection is not apparent. The selection for peptides with Glu at the +2 position by Src family SH2 domains is thought to be at least partially influenced by the Arg residue at the β D'1 position of the SH2 domain (45, 49). The SH2 domains of Abl, GRB-2 members, and 3BP2 have distinct amino acids at the β D'1 position (Table 2). It seems likely that the very strong selection for Asn at the +2 position by GRB-2 family members is due to a pocket being formed in the SH2 domain at this location. Crystal or solution nuclear magnetic resonance determination of the structures of these proteins will be necessary to elucidate this selection.

The motifs selected by the group III SH2 domain proteins were also consistent with previous predictions. This group of SH2 domains has Ile or Cys at the β D5 position and selects phosphopeptides on the basis of hydrophobic amino acids at the +1 and +3 positions carboxy terminal of the phospho-Tyr (45). Unlike group I members, they exhibit very little selection on the basis of the +2 position (45). The peptides selected by the amino-terminal SH2 domain of HCP certainly fit these criteria. The SHC SH2 domain has a Leu at the β D5 position rather than Ile or Cys but seems to fit in group III on the basis of selecting the motif phospho-Tyr-Glu/Ile-Xxx-Ile/Leu/Met.

The Vav SH2 domain has a Thr at the β D5 position, placing it in a group by itself (group II). This SH2 domain selects phosphopeptides with Met/Leu/Glu at the +1 position. The selection at this position is similar to that of the N-SH2 domain of the p85 subunits of PI-3-kinase (Met/Val/Ile/Glu [Table 2]). The p85 SH2 domain has a Lys at β D3 and an Ile at β D5 (the two positions predicted to influence specificity at the +1

TABLE 2. Phosphopeptide motifs for SH2 domains: residues predicted to interact with the side chains of the associated phosphopeptides^a

Group	SH2 domain ^e	Src-SH2			Src-SH2		Src-SH2					
		+1	200 βD3	202 βD5	+2	205 βD'1	+3	202 βD5	214 βE4	215 EF1	230 αB9	237 BG4
1A	Src	E	K	Y	E	R	I	Y	I	T	Y	L
	Fyn	E	K	Y	E	R	I	Y	I	T	Y	L
	Lck	E	K	Y	E	R	I	Y	I	S	Y	L
	Fgr	E	K	Y	E	R	IV	Y	I	T	Y	L
	Lyn		K	Y		R		Y	I	S	Y	L
	Yes		K	Y		R		Y	I	T	Y	L
	Hck		K	Y		R		Y	I	S	Y	L
	Dsrc		K	Y		K		Y	L	S	Y	L
1B	Syk N ^b		H	Y		E		Y	I	S	H	L
	*Syk C ^c	QTE	L	Y	eqt	D	L	Y	I	P	Y	L
	ZAP70 C		Y	Y		S		Y	I	P	L	L
	Tec		R	Y		K		Y	L	A	H	L
	Atk		R	Y		C		Y	L	A	H	L
	Itk		K	Y		K		Y	V	A	H	L
	Abl	E	Y	Y	N	N	P	Y	V	S	H	L
	Arg		Y	Y		N		Y	V	T	H	L
	*Csk	T	E	Y	K	M	MIVr	Y	I	D	Y	?
	Crk	D	S	Y	H	N	P	Y	A	G	Y	T?
	Nck	D	K	F	E	Q	P	F	I	G	Y	T?
	*fps/fes	E	R	F	—	Q	vi	F	R	L	L	G
	ZAP70 N		H	F		E		F	I	A	Y	L
	Sem5	LV	Q	F	N	L	vp	F	L	W	H	R?
	*DGBR2	y	Q	F	N	L	—	F	L	W	H	R?
	*GRB2	qy	Q	F	N	L	y	F	L	W	H	R?
	GAP ^d C		Q	F		C		F	M	G	Y	I?
	GAP N		N	F		I		F	I	G	Y	L?
	Tensin		R	F		T		F	?	?	H	?
	*3BP2	E	R	Y	N	F	—	Y	E	G	Y	P?
2	*Vav	M	K	T	E	I	P	T	I	T	Y	?
3	p85aN	MIVE	K	I	—	F	M	I	F	S	Y	A?
	p85bN		K	I		F		I	F	S	Y	A?
	p85aC	mli	K	C	—	N	M	C	F	A	Y	V?
	p85bC		K	C		Y		C	F	A	Y	V?
	PLC-γ 1C	VI	K	C	IL	N	PIV	C	L	G	Y	Y?
	PLC-γ 2C		K	C		Q		C	L	G	Y	Y?
	PLC-γ 1N	LIV	Q	C	Ed	H	LIV	C	K	F	Y	L?
	PLC-γ 2N		Q	C		R		C	K	Y	Y	L?
	*HCP N	F	T	I	—	Q	Fply	I	D	L	Y	V
	SHPTP2 N	IV	T	I	—	Q	VI	I	D	L	Y	L
	CSW N		T	I		Q		I	D	L	Y	L
	SHPTP1 C		T	I		M		I	T	V	F	E
4	*SHC	EI	K	L	—	V	ILM	L	T	K	H	P?
	ShB		M	M		A		M	L	G	Y	?
	SHPTP2 C		T	V		R		V	D	V	Y	E
	CSW C		T	V		R		V	D	V	Y	E
	113 TF		F	A		P		A	L?	S?	Y	?
	91 TF		I	S		P		S	L?	Q?	Y	?

^a Columns +1, +2, and +3 comprise the first, second, and third residues C-terminal to phospho-Tyr of the optimal phosphopeptide selected by each SH2 domain (e.g., P-YEEI for Src SH2). Src SH2 200 and 202 indicate the residues of Src (and residues at analogous positions of other SH2 domains) predicted to contact the +1 residue side chain of the associate peptide. Src SH2 205 is predicted to be near the +2 side chain, and Src 202, 214, 215, and 237 are predicted to form a hydrophobic pocket to bind the +3 residue side chain. The alignments were made on the basis of data from Waksman et al. (49). Boldface letters indicate strong selection, uppercase nonboldface letters indicate medium selection, and lowercase letters indicates weak selection (—, no selection). Motifs not yet determined or not submitted for publication are left blank.

^b N, N terminal.

^c C, C terminal.

^d GAP, GTPase-activating protein.

^e Asterisks indicate SH2 domains studied in this work.

residue of the associated phosphopeptide). Vav also has a Lys at βD3 but, as indicated above, has Thr rather than Ile at βD5. The fact that these two SH2 domains select peptides with similar residues at the +1 position suggests that the branched

methyl groups at the beta-carbons common to the βD5 Ile and βD5 Thr of these two SH2 domains may be making similar contacts with the +1 side chains of the associated peptides.

Potential in vivo binding sites for the SH2 domains. On the

SH2 Domain	Motif	Binding Sites in Proteins
SH2	PO4 Y T K M A R I S Q V N R	PO4 Y M X V I
SH2	POTENTIAL SITES	POTENTIAL SITES
DTSSVLYTAVQP	Hum PDGF Receptor β	HUM IGF1 RECEPTOR (Y1014)
LLFEE YTN1 PI	Human HTLV I TAT protein	CHICK FAK (Y901)
IEDNE YTAQ OG	src Tyrosine Kinase	MOUSE TEC TYR KIN. (Y468)
IEDNE YTAQ OG	HUM YES	HUM LYN TYR KIN. (Y434)
IEDNE YTAQ OG	Hum fyn	HUM C-ABL TYR KIN. (Y257)
IEDNE YTAQ OG	Hum hck	HUM HCK TYR KIN (Y68)
IEDNE YTAQ OG	Hum lyn	MOUSE STY KINASE (Y160)
IEDNE YTAQ OG	Hum lck	HUM PLSTIRE S/T PROT. KIN. (Y13)
LNFEK YTQM LN	TURKEY C-REL PROTO-ONCOGENE	HUM B-CELL DIFF. ANTIGEN CD72 (Y39)
YGERP YSQM SN	MOUSE MKK4 RECEPTOR TYR KINASE	MOUSE T-CELL CD3 EPSILON (Y170)
		MOUSE HIGH AFFINITY IGF FC RCPTR (Y357)
		MUS GCSF RECEPTOR (Y788)
		MOUSE HCP/SHPTPI (Y564)
		HUM LAR TRANSMEMB. PTPASE (Y1346)
		HUM EZRIN (Y477)
		BCR HOMOLOGUE (Y42)
		MUS 3BP2a (Y173)
		MOUSE GAMMA ADAPTIN (Y289)
3BP2 SH2	PO4 Y M X M V V I	PO4 Y F X F P L Y
	POTENTIAL SITES	POTENTIAL SITES
EDIEE YENQ KR	MOUSE FLK-2 RECEPTOR TYR KIN.	HUM PIP2 phosphodiesterase (Y292)
QARTO YENF IH	HUM IGF-1 RECEPTOR.	HUM RCP Receptor (Y992)
SNOKV YENF TGLVK	CHICK FAK	HUM ras-GAP (Y509)
EGKIP YENR SNSEV	MOUSE ITK TYR KINASE	HUM TYR2 TYROSINE KINASE (Y494)
EDSEF YEND SNLQO	HUM B CELL CD19	HUM TYK2 TYROSINE KINASE (Y423)
QDQSO YENF EDEFL	*	HUM PTPase-meg (Y69)
SNASE YENE DEELT	*	NFFAH YPTL GS (Y324)
EDADS YENM DNPDG	*	SLVDC YFRL TA (Y400)
ARPCA YENC RADLQ	MOUSE LIMP.-PEYERS PATCH A4	ABRLI YPOF YN (Y975)
GFLPK YENK TYLAV	*	ISNTD YFFP R (Y464)
YENPN YENM LIPAA	HUM EPO RECEPTOR	PTNQC YFFP HL (Y364)
ALKYM YENH NISIT	HUM HIGH AFFINITY EPSILON RECEPTOR	EACOV YPTD DP (Y381)
PSPKS YENL WFOAS	HUM GCSF3_1	PLQLQ YPSP YM (Y226)
DGEIT YENM QVPAV	HUM B-CELL DIFF. ANTIGEN CD72	NDHIC YFLV QI (Y145)
ARLEY YENE KXWHH	MOUSE IRS1_1	HUM DBL PROTOONCOGENE (Y899)
HKEDV YENL HTNKK	HUM HCP/SHPTPI	
EDIEE YENQ KRLAE	MOUSE FLT3	
	PO4 Y Q E L T Q I E	PO4 Y M E P L E
	POTENTIAL SITES	POTENTIAL SITES
EDDHT YEGL NIDQAT YEDI VT	Mouse B-cell B29 antigen	HUM PDGF Receptor β (Y771)
EDENL YEGL NLDDCSM YEDI SR	Mouse B-cell IgM-assoc. Mb-1	VDHDE YLIF QO (Y992)
KPESD YQAL LPSAPEI YSHL SP	Bovine leukaemia vir. gp30	IGAPG YMEC SA (Y156)
KADAV YTGL NTRSQRT YETL KH	Mouse Ig epsilon Recept g	COTIE YMAP EI (Y259)
PDRRL YEEL NVYSPI YSEL ED	Mouse Ig epsilon Recept b	GAREP YMEC VN (Y952)
BDEIE YEDL MR	Human VAV oncogene	HLGLE YHTF WL (Y1594)
LLFEE YTN1 PI	Human HTLV I TAT protein	RAT ERK3 KINASE (Y110)
LLSKE YEEL KD	Human Protein-Tyr Phosphatase b	HUM ECK TYROSINE KINASE (Y694)
SFSTI YOEL OS	Human fcs/fea tyrosine kinase	WHOGI YMEP EN (Y416)
		PULSR YMED ST (Y545)
		MYITE YMEN GS (Y701)
		YIITE YMEN GS (Y318)
		EBEGB YEFP DS (Y409)
		ELVQH YMEV ND (Y226)
		AEQMA YMEP MN (Y372)
		EEOTE YHTP SS (Y700)
	PO4 Y Q M Y Y Q V F	PO4 Y M F I L V I V
	POTENTIAL SITES	POTENTIAL SITES
GRB2 SH2	HUM EGF Receptor (Y1092)	HUM EGF Receptor (Y1092)
LPVPE YINQ SV	*	HUM EGF Receptor (Y1092)
VGNPE YLNT VQ	HUM SHW (Y1138)	HUM EGF Receptor (Y1092)
PDPRS YNVN QN	C. elegans LET-23 tyr kin (Y117)	HUM EGF Receptor (Y1092)
PSNGID YVNO PN	S. elegans LET-23 tyr kin (Y1276)	HUM EGF Receptor (Y1092)
SSSSG YVNE PH	Drosophila son of sevenless protein (Y1289)	HUM EGF Receptor (Y1092)
GEIQQ YVNO PY	Drosophila son of sevenless protein (Y1020)	HUM EGF Receptor (Y1092)
	POTENTIAL SITES	POTENTIAL SITES
HVNAT YVNV LC	Hum HGF Receptor (Y1374)	HUM EGF Receptor (Y1092)
SPQPE YVNO PD	Hum Erb B2 (Y1139)	HUM EGF Receptor (Y1092)
ADGMA YVNA NK	Hum IGF-1 recept. (Y1125)	HUM EGF Receptor (Y1092)
ADGMA YVNA KK	Hum Insulin Recept. (Y4149)	HUM EGF Receptor (Y1092)
SDOVR YVNA FK	Hum flt tyrosine kinase (Y1213)	HUM EGF Receptor (Y1092)
DAGSN YVNA SY	Hum CD45 PTPase (Y706)	HUM EGF Receptor (Y1092)
EPSPK YVNA SF	*	HUM EGF Receptor (Y1092)
LDGHE YVNO PQ	platyfish melanoma receptor protein (Y1015)	HUM EGF Receptor (Y1092)
KEOPH YVNO TQ	Drosophila kraf (Y1057)	HUM EGF Receptor (Y1092)
TMSPF YVNO	Human interleukin-7 receptor (Y456)	HUM EGF Receptor (Y1092)
	PO4 Y I X I E L Y M	PO4 Y M F I L V I V
	POTENTIAL SITES	POTENTIAL SITES
SHC SH2	Human Erb B3 (Y1270)	HUM EGF Receptor (Y1092)
ASEQG YEEM RA	*	HUM EGF Receptor (Y1092)
TPDED YEYM NR	*	HUM EGF Receptor (Y1092)
DEDER YEYM NR	*	HUM EGF Receptor (Y1092)
LEELG YEYM DV	*	HUM EGF Receptor (Y1092)
EELS N YICM GG	Rat Insulin Rec. Substrate 1 (Y460)	HUM EGF Receptor (Y1092)
AEENP YLGL DV	Human ERB B2 (Y1248)	HUM EGF Receptor (Y1092)
IBNPQ YFQI TN	X17647 Mouse trkb-1 bdnf receptor (Y515)	HUM EGF Receptor (Y1092)
YVNN YEDL TD	platyfish melanoma receptor protein (Y1077)	HUM EGF Receptor (Y1092)
LNENP YVSM RS	Mouse Polyoma MT (Y250)	HUM EGF Receptor (Y1092)
EEEPQ YEEI PI	Hamster Polyoma MT (Y324)	HUM EGF Receptor (Y1092)
QONQL YNEL NLGRREE YDVL DK	Human CD3 zeta chain (Y72, Y83)	HUM EGF Receptor (Y1092)
GHIDL YQGL STATROT YDAL RH	*	HUM EGF Receptor (Y1092)
PORCL YNEL QKQMAEA YSEI QH	*	HUM EGF Receptor (Y1092)
PMDQL YQPL KDREDDQ YSHL QG	Human CD3 gamma chain (Y110, Y122)	HUM EGF Receptor (Y1092)
RNDQV YQPL RDREDDQ YSHL QG	Human CD3 delta chain (Y160, Y171)	HUM EGF Receptor (Y1092)
VNPD YEP1 RKQQRDL YSLG NQ	Human CD3 epsilon chain (Y149, Y160)	HUM EGF Receptor (Y1092)
EDDHT YEGL NIDQAT YEDI VT	Mouse B-cell B29 antigen (Y170, Y181)	HUM EGF Receptor (Y1092)
KADAV YTGL NTRSQRT YETL KH	Mouse Ig epsilon Recept g (Y195, Y206)	HUM EGF Receptor (Y1092)
PDRRL YEEL NVYSPI YSEL ED	Mouse Ig epsilon Recept b (Y65, Y76)	HUM EGF Receptor (Y1092)
EDENL YEGL NLDDCSM YEDI SR	Mouse B-cell IgM-assoc. Mb-1 (Y210, Y220)	HUM EGF Receptor (Y1092)
BDEIE YEDL MR	Human VAV oncogene (Y182, Y193)	HUM EGF Receptor (Y1092)
AAACV YEDM SH	Human T-cell CD7 (Y126)	HUM EGF Receptor (Y1092)
LTSEE YEEL RG	Bovine Polyoma Large T (Y222)	HUM EGF Receptor (Y1092)
NLSKR YEEI YL	Human retinoblastoma-assoc. rb 110 (Y8)	HUM EGF Receptor (Y1092)
YDNA YEDI DV	Hum papillomavirus (type 16) L2 protein (Y321)	HUM EGF Receptor (Y1092)
LLSKE YEEL KD	Human Protein-Tyr Phosphatase b (Y1707)	HUM EGF Receptor (Y1092)
RRERD YTNL PS	Human CSF-1 receptor (Y923)	HUM EGF Receptor (Y1092)

basis of the preferred motifs, we have scanned the protein data base for likely in vivo targets of the various SH2 domains. Some potential sites in proteins known to be involved in signal transduction are summarized in Fig. 2. These proteins were

FIG. 2. Phospho-Tyr peptide specificity of SH2 domains. Sequences with similarity to the motifs that bind the indicated SH2 domains are presented. The location of Tyr in the primary sequence is indicated in parentheses. The sequences were found with the Blast, FASTA, and FIND programs used to search the GenBank data base. A subgroup of proteins potentially involved in signaling where the SH2-binding motif is found in a cytosolic domain carboxy terminal to a negatively charged region (likely phosphorylation site) is presented. Abbreviations: Hum and HUM, human; Ig and IG, immunoglobulin; MT, middle T antigen; vir., virus; Rec. and RECPTR, receptor; KIN., kinase; DIFF., differentiation; TRANSMEMB., transmembrane; as-soc., associated; HTLV I, human T-cell lymphotropic virus type I; EPO, erythropoietin; EGF, epidermal growth factor; CSF-1, colony-stimulating factor 1; GCSF, granulocyte colony-stimulating factor; IL, interleukin; PTPASE, protein tyrosine phosphatase; GAP, GTPase-activating protein.

identified with the Genetics Computer Group FIND PATTERN program by using the optimal sequences predicted by the phosphopeptide library. The search was restricted to receptors, protein kinases, and protein kinase substrates. Of course, most of the sites identified in Fig. 2 have not yet been investigated for the possibility of in vivo phosphorylation. However, the predictions made in Fig. 2 can easily be tested in cases in which the precipitating antibodies are available. Ultimately, if a protein identified in Fig. 2 is shown to coimmunoprecipitate with the SH2-containing protein, the tyrosine residue of the putative motif can be mutated to test the predicted binding site. Thus, the laborious step of mapping

the *in vivo* phosphorylation sites can be circumvented. We are encouraged by the number of predicted SH2-binding sites (4, 45) that have already been confirmed (11, 20, 22, 32, 34). Some of the predictions in Fig. 2 suggest interesting models for the involvement of these proteins in signaling cascades.

Csk SH2 domain. The motif determined for the Csk SH2 domain suggests a mechanism by which this kinase targets Src family members *in vivo*. pp60^{c-src} has two major sites of phosphorylation *in vivo*. Autophosphorylation at Tyr-416 causes an increase in the activity of pp60^{c-src} toward exogenous substrates (13). Csk was isolated because of its ability to phosphorylate Tyr-527 of pp60^{c-src} (30). Phosphorylation at Tyr-527 results in inhibition of pp60^{c-src} tyrosine kinase activity via a head-to-tail association with the SH2 domain of the same molecule (4). As indicated in Fig. 2, the sequence C terminal of the Tyr-416 autophosphorylation site in pp60^{c-src} is in good agreement with the motif selected by the Csk SH2 domain. This result suggests that Csk binds to autophosphorylated (activated) pp60^{c-src} via its SH2 domain and then phosphorylates Tyr-527 to shut off the response. Other members of the Src family, except Fgr, have a conserved Tyr-Thr-Ala-Arg motif at the position homologous to Tyr-416 of pp60^{c-src}, suggesting that they all interact with Csk in a similar fashion (Fig. 2). Fgr is interesting in that the analogous sequence at the autophosphorylation site is Tyr-Asn-Pro-Cys. This sequence is similar to the motif at the C-terminal Tyr of Fgr (Tyr-Gln-Pro-Gly), suggesting that the autophosphorylation site might compete with the C-terminal phosphorylation site for binding to the SH2 domain of the same molecule or the sister molecule of a homodimer. Consistent with this idea, autophosphorylated Fgr is resistant to inhibition by Csk, even though the C-terminal Tyr can be phosphorylated (42).

The selections for Thr at +1 and Arg at +3 are unique to Csk of all the SH2 domains investigated to date. An Asp residue at position EF1 that is predicted to form the +3 pocket (unique to Csk [Fig. 2]) may account for the unusual selection for Arg. The Glu at β D3 is also unique to Csk and is likely to contribute to the selection for Thr at +1.

3BP2 SH2 domain. 3BP2 was isolated as a protein that binds tightly to the SH3 domain of Abl (37). In addition to the proline-rich region that mediates SH3 binding, it has a pleckstrin homology domain (25) and an SH2 domain. As can be seen in Fig. 2, a number of receptors, tyrosine kinases, and tyrosine kinase substrates have the Tyr-Glu-Asn-Xxx motif optimal for the 3BP2 SH2 domain. B-cell CD19 has four closely spaced copies of this motif.

Syk C-SH2 domain. The Syk and ZAP70 protein tyrosine kinases are expressed in B cells and T cells, respectively, and are implicated in signaling by directly binding to B- and T-cell receptor-associated proteins (6, 46). They make up a novel family of cytosolic protein tyrosine kinases with tandem SH2 domains at the amino terminus. We previously speculated that the two SH2 domains of these proteins might simultaneously bind to two closely spaced phospho-Tyr residues on T-cell (or B-cell) receptor-associated proteins to provide a high-affinity bidentate contact (45). The argument for this model stems from the observation that the spacing between Tyr residues in these receptor-associated proteins [Tyr-Xxx-Xxx-Leu/Ile-(Xxx)-Tyr-Xxx-Xxx-Leu/Ile (38)] is identical to the spacing between the Tyr-Xxx-Xxx-Met motifs in the PDGF receptor that bind the SH2 domains of the PI-3-kinase. Doubly phosphorylated peptides, on the basis of these motifs, are much more effective than singly phosphorylated peptides in activating PI-3-kinase, suggesting simultaneous binding to both SH2 domains of the p85 subunit (5). We predicted that the SH2 domains of Syk and ZAP70 would select phospho-Tyr-hydro-

philic-hydrophilic-Leu/Ile motifs (consistent with the Reth motifs) on the basis of their similarity to Src family SH2 domains (45). The results with the Syk C-SH2 domain confirm this prediction (Fig. 2). In addition to the B-cell-specific proteins with Reth motifs, the Vav oncoprotein, human T-cell leukemia virus I TAT protein, protein tyrosine phosphatase B, and fps/fes protein tyrosine kinase also have potential binding sites for Syk (Fig. 2). Attempts to determine motifs for the N-SH2 domain of Syk and the SH2 domains of ZAP70 have thus far failed because of problems in the expression of sufficient functional protein in bacteria.

fps/fes SH2 domain. *fps/fes* was isolated as a retrovirus-encoded oncogene (15). It is a distant relative of the Src family. Its SH2 domain strongly selects for phosphopeptides with Glu at +1. There may be overlap with targets selected by Src family members, but unlike the Src family SH2 domains, the *fps/fes* SH2 domain prefers peptides with Val at the +3 position. Possible sites are shown in Fig. 2.

HCP N-SH2 domain. HCP is a protein tyrosine phosphatase expressed in lymphocytes (52). Like the N-SH2 domain of the close homolog SHPTP2/syp, the N-SH2 domain of HCP has relatively broad selectivity for sites with hydrophobic residues +1 and +3 of phospho-Tyr. The preferred hydrophobic residue at +1 and +3 is Phe. Of the potential sites shown in Fig. 2, those on the interleukin 2 and 6 receptors and high-affinity immunoglobulin Fc receptor are predicted to have the highest affinity.

GRB-2/DGRB-2/Drk SH2 domain. GRB-2, Drk, and Sem5 are homologs of SH2- and SH3-containing proteins from humans, *D. melanogaster*, and *C. elegans* that mediate receptor coupling to the *ras* activator, SOS (32). Consistent with the motif selected by the Sem5 SH2 domain, GRB-2 and DGRB-2/Drk select phosphopeptides primarily on the basis of the presence of an Asn at +2 (as discussed above). Of the sites in Fig. 2 predicted to bind GRB-2, Tyr-317 of SHC and Tyr-1092 of the epidermal growth factor (EGF) receptor have already been shown to mediate GRB-2 binding (32).

SHC SH2 domain. As indicated in the Results section, the motif selected by the SHC SH2 domain (phospho-Tyr-Ile/Glu/Tyr-Xxx-Ile/Leu/Met) suggests that it could bind to the Tyr-Xxx-Xxx-Leu/Ile motifs that are repeated on subunits of the T-cell receptor CD3 complex and the B-cell receptor-associated proteins (38). The possibility that SHC binds to a component of CD3 was investigated (35). SHC was shown to bind to CD3 ξ in response to T-cell receptor stimulation, and this association occurred through the SH2 domain and was blocked by a phosphopeptide on the basis of Tyr-Xxx-Xxx-Leu motif in CD3 ξ (35). These results provide a mechanism for T-cell receptor activation of *ras* via assembly of a CD3 ξ /SHC/GRB-2/SOS complex. The multiple Tyr-Xxx-Xxx-Leu/Ile motifs in T- and B-cell receptors are likely to bind to multiple SH2-containing proteins (see discussion of the Syk SH2 domain below). Those sites with Ile at the +3 position are predicted to bind either SHC or Src family members (Lck, Fyn, Lyn, etc.) (45).

The selection for phosphopeptides with Met at the +3 position indicates that SHC might compete at some sites that bind p85 of PI-3-kinase (45). Because of the stronger selection for Met at the +3 position, p85 is expected to bind much more tightly at these sites, yet some of the multiple Tyr-Glu/Ile-Xxx-Met sites on IRS-1 and ERB B3 are good candidates for SHC binding sites (Fig. 2).

Interestingly, a number of the candidate sites for binding SHC SH2 domains have the sequence Asn-Pro-Xxx amino terminal of the Tyr residue (29). This motif is found amino terminal of the sites predicted in ERB B2, trkb-1 (bdfn

receptor), melanoma receptor protein, mouse polyomavirus middle T antigen, and mouse CD3 ϵ . Mutations of the Asn-248, Pro-249, or Tyr-250 of polyomavirus middle T antigen have been shown to impair transformation, suggesting that these residues function together as a motif (8, 24). In addition, the site on the nerve growth factor receptor mapped for SHC binding is preceded by Asn-Pro-Xxx (15a). The Asn-Pro-Xxx-Tyr/Phe motif was previously pointed out as a sequence necessary for low-density lipoprotein receptor localization in coated pits (7). This sequence, when combined with hydrophobic residues upstream, is predicted to form a bend (2). Such a bend may be important for presenting the phospho-Tyr motif at the surface of the protein for interaction with the SHC SH2 domain. Alternatively, the amino-terminal residues may make direct contact with the SHC SH2 domain. We are currently exploring the possibility that certain SH2 domains recognize residues amino terminal of the phospho-Tyr residue (28) by using a library with degeneracy in these residues. Thus far, most of the SH2 domains investigated showed little selection for residues amino terminal of phospho-Tyr (44a).

Vav SH2 domain. *vav* was isolated as a human oncogene (3). In addition to the SH2 domain, it has a domain with weak homology to *dbf* (39), another human oncogene. It is normally expressed in lymphocytes, and it was recently implicated in the regulation of p21^{ras} (14). Although the in vivo targets of the Vav SH2 domain are still not clear, previous reports have indicated that this protein can bind to the ligand-stimulated EGF- and PDGF-receptors (3). In agreement with these biological data, both the EGF and PDGF receptors contain potential docking sites for Vav (Fig. 2).

In summary, the motifs determined for the various SH2 domains offer explanations for specificity in protein tyrosine kinase signaling cascades. They also make testable predictions for in vivo interactions on the basis of the primary sequences in the protein data base. These predictions are likely to accelerate progress in understanding signaling cascades.

ACKNOWLEDGMENTS

We thank Michael Berne (Tufts University Protein Chemistry) for sequencing the peptide mixtures.

This research was supported by a grant from the Lucille P. Markey Charitable Trust (L.C.C.), by the BASF Bioresearch Corporation, and by grants from the NIH.

REFERENCES

- Anderson, D., C. A. Koch, L. Grey, C. Ellis, M. F. Moran, and T. Pawson. 1990. Binding of SH2 domains of phospholipase C gamma 1, GAP, and Src to activated growth factor receptors. *Science* **250**:979-982.
- Bansal, A., and L. M. Gierasch. 1991. The NPXY internalization signal of the LDL receptor adopts a reverse-turn conformation. *Cell* **67**:1195-1201.
- Bustelo, X. R., A. J. Ledbetter, and M. Barbacid. 1992. Product of Vav proto-oncogene defines a new class of tyrosine protein kinase substrates. *Nature (London)* **356**:68-71.
- Cantley, L. C., K. R. Auger, C. L. Carpenter, B. Duckworth, A. Graziani, R. Kapeller, and S. Soltoff. 1991. Oncogenes and signal transduction. *Cell* **64**:281-302.
- Carpenter, C. L., K. R. Auger, M. Chaudhuri, M. Yoakim, B. Schaffhausen, S. Shoelson, and L. C. Cantley. 1993. Phosphoinositide 3-kinase is activated by phosphopeptides that bind to the SH2 domains of the 85-kDa subunit. *J. Biol. Chem.* **268**:9478-9483.
- Chan, A. C., B. A. Irving, J. D. Fraser, and A. Weiss. 1991. The zeta chain is associated with a tyrosine kinase and upon T-cell antigen receptor stimulation associates with ZAP-70, a 70-kDa tyrosine phosphoprotein. *Proc. Natl. Acad. Sci. USA* **88**:9166-9170.
- Chen, W.-J., J. L. Goldstein, and M. S. Brown. 1990. NPXY, a sequence often found in cytoplasmic tails, is required for coated pit-mediated internalization of the low density lipoprotein receptor. *J. Biol. Chem.* **265**:3116-3128.
- Druker, B. J., L. E. Ling, B. Cohen, T. M. Roberts, and B. S. Schaffhausen. 1990. A completely transformation-defective point mutant of polyomavirus middle T antigen which retains full associated phosphatidylinositol kinase activity. *J. Virol.* **64**:4454-4461.
- Eck, M. J., S. E. Shoelson, and S. C. Harrison. 1993. Recognition of a high-affinity phosphotyrosyl peptide by the Src homology-2 domain of p56lck. *Nature (London)* **362**:87-91.
- Escobedo, J. A., D. R. Kaplan, W. M. Kavanaugh, C. W. Turck, and L. T. Williams. 1991. A phosphatidylinositol-3 kinase binds to platelet-derived growth factor receptors through a specific receptor sequence containing phosphotyrosine. *Mol. Cell. Biol.* **11**:1125-1132.
- Fantl, W. J., J. A. Escobedo, G. A. Martin, C. W. Turck, R. M. Del, F. McCormick, and L. T. Williams. 1992. Distinct phosphotyrosines on a growth factor receptor bind to specific molecules that mediate different signaling pathways. *Cell* **69**:413-423.
- Feldman, R. A., W. C. Vass, and P. E. Tambourin. 1987. Human cellular fps/fes cDNA rescued via retroviral shuttle vector encodes myeloid cell NCP92 and has transforming potential. *Oncogene Res.* **1**:441-458.
- Fukami, Y., K. Sato, K. Ikeda, K. Kamisano, K. Koizumi, and A. T. Matsuno. 1993. Evidence for autoinhibitory regulation of the c-Src gene product. *J. Biol. Chem.* **268**:1132-1140.
- Gulbins, E., K. M. Coggeshall, G. Baier, S. Katrav, P. Burn, and A. Altman. 1993. Tyrosine kinase-stimulated guanine nucleotide exchange activity of Vav in T cell activation. *Science* **260**:822-825.
- Hanafusa, H. 1988. The fps/fes oncogene, p. 39-57. In E. P. Reddy, A. M. Skalka, and T. Curran (ed.), *The oncogene handbook*. Elsevier Science Publishers, New York.
- 15a. Kaplan, D. Unpublished result.
- Kaplan, D. R., D. K. Morrison, G. Wong, F. McCormick, and L. T. Williams. 1990. PDGF beta-receptor stimulates tyrosine phosphorylation of GAP and association of GAP with a signaling complex. *Cell* **61**:125-133.
- Kaplan, D. R., M. Whitman, B. Schaffhausen, D. C. Pallas, M. White, L. Cantley, and T. M. Roberts. 1987. Common elements in growth factor stimulation and oncogenic transformation: 85 kd phosphoprotein and phosphatidylinositol kinase activity. *Cell* **50**:1021-1029.
- Katzav, S., D. Martin-Zanca, and M. Barbacid. 1989. Vav, a novel human oncogene derived from a locus ubiquitously expressed in hematopoietic cells. *EMBO J.* **8**:2283-2290.
- Kazlauskas, A., and J. A. Cooper. 1989. Autophosphorylation of the PDGF receptor in the kinase insert region regulates interactions with cell proteins. *Cell* **58**:1121-1133.
- Kazlauskas, A., G.-S. Feng, T. Pawson, and M. Valius. 1993. The 64-kDa protein that associates with the platelet derived growth factor β subunit via Tyr-1009 is the SH2-containing phosphotyrosine phosphatase Syp. *Proc. Natl. Acad. Sci. USA* **90**:6939-6942.
- Kypta, R. M., Y. Goldberg, E. T. Ulug, and S. A. Courtneidge. 1990. Association between the PDGF receptor and members of the src family of tyrosine kinases. *Cell* **62**:481-492.
- Lechleider, R. J., S. Sugimoto, A. M. Bennet, J. A. Cooper, S. E. Shoelson, C. T. Walsh, and B. G. Neel. 1993. Activation of the SH2-containing phosphotyrosine phosphatase SH-PTP2 by its binding site, phosphotyrosine 1009 on the PDGF receptor. *J. Biol. Chem.* **268**:21478-21481.
- Lev, S., D. Givol, and Y. Yarden. 1992. Interkinase domain of kit contains the binding site for phosphatidylinositol 3' kinase. *Proc. Natl. Acad. Sci. USA* **89**:678-682.
- Ling, L. E., B. J. Druker, L. C. Cantley, and T. M. Roberts. 1992. Transformation-defective mutants of polyomavirus middle T antigen associate with phosphatidylinositol 3-kinase (PI 3-kinase) but are unable to maintain wild-type levels of PI 3-kinase products in intact cells. *J. Virol.* **66**:1702-1708.
- Mayer, B. J., R. Ren, K. L. Clark, and D. Baltimore. 1993. A putative modular domain present in diverse signaling proteins. *Cell* **73**:629-630.

26. McGlade, C. J., C. Ellis, M. Reedijk, D. Anderson, G. Mbamalu, A. D. Reith, G. Panayotou, P. End, A. Bernstein, A. Kazlauskas, M. D. Waterfield, and T. Pawson. 1992. SH2 domains of p85 α subunit of phosphatidylinositol 3-kinase regulate binding to growth factor receptors. *Mol. Cell. Biol.* **12**:991-997.
27. Meisenhelder, J., P. G. Suh, S. Rhee, and T. Hunter. 1989. Phospholipase C-gamma is a substrate for the PDGF and EGF receptor protein-tyrosine kinases in vivo and in vitro. *Cell* **57**:1109-1122.
28. Nishimura, R., W. Li, A. Kashishian, A. Mondino, M. Zhou, J. Cooper, and J. Schlessinger. 1993. Two signaling molecules share a phosphotyrosine-containing binding site in the platelet-derived growth factor receptor. *Mol. Cell. Biol.* **13**:6889-6896.
29. Obermeier, A., R. Lammers, K. Wiesmuller, G. Jung, J. Schlessinger, and A. Ullrich. 1993. Identification of Trk binding sites for SHC and phosphatidylinositol 3'-kinase and formation of a multimeric signaling complex. *J. Biol. Chem.* **268**:22963-22966.
30. Okada, M., and H. Nakagawa. 1989. A protein tyrosine kinase involved in regulation of pp60^{c-src} function. *J. Biol. Chem.* **264**:20886-20893.
31. Olivier, J. P., T. Raabe, M. Henkemeyer, B. Dickson, G. Mbamalu, B. Margolis, J. Schlessinger, E. Hafen, and T. Pawson. 1993. A Drosophila SH2-SH3 adaptor protein implicated in coupling the sevenless tyrosine kinase to an activator of Ras guanine nucleotide exchange, Sos. *Cell* **73**:179-191.
32. Pawson, T., and J. Schlessinger. 1993. SH2 and SH3 domains. *Curr. Biol.* **3**:434-442.
33. Pelicci, G., L. Lanfrancone, F. Grignani, J. McGlade, F. Cavallo, G. Forni, I. Nicoletti, F. Grignani, T. Pawson, and P. G. Pelicci. 1992. A novel transforming protein (SHC) with an SH2 domain is implicated in mitogenic signal transduction. *Cell* **70**:93-104.
34. Prasad, K. V. S. (Harvard Medical School). Personal communication.
35. Ravichandran, K. S., K. K. Lee, Z. Songyang, L. C. Cantley, P. Burn, and S. J. Burakoff. 1993. SHC interacts with the Zeta chain of the T cell receptor upon T cell activation. *Science* **262**:902-905.
36. Reedijk, M., X. Liu, P. V. D. Geer, K. Letwin, M. D. Waterfield, T. Hunter, and T. Pawson. 1992. Tyr 721 regulates specific binding of the CSF-1 receptor kinase insert to PI3-kinase SH2 domains: a model for SH2-mediated receptor target interactions. *EMBO J.* **11**:1365-1372.
37. Ren, R., B. J. Mayer, P. Cicchetti, and D. Baltimore. 1993. Identification of a ten-amino acid proline-rich SH3 binding site. *Science* **259**:1157-1161.
38. Reth, M. 1989. Antigen receptor tail clue. *Nature (London)* **338**:383-384.
39. Ron, D., S. R. Tronick, S. A. Aaronson, and A. Eva. 1988. Molecular cloning and characterization of the human dbl proto-oncogene: evidence that its overexpression is sufficient to transform NIH/3T3 cells. *EMBO J.* **7**:2465-2473.
40. Ronnstrand, L., S. Mori, A. Arridsson, A. Eriksson, C. Wernstedt, C. Hellman, L. Claesson-Welsh, and C. Heldin. 1992. Identification of two C-terminal autophosphorylation sites in the PDGF β receptor: involvement in the interaction with phospholipase C- γ . *EMBO J.* **11**:3911-3919.
41. Rotin, D., B. Margolis, M. Mohammadi, R. Daly, G. Drum, N. Li, E. Fisher, W. Burgess, A. Ullrich, and J. Schlessinger. 1992. SH2 domains prevent tyrosine dephosphorylation of the EGF receptor: identification of tyr 992 as the high-affinity binding site for SH2 domains of phospholipase C-gamma 1. *EMBO J.* **11**:559-567.
42. Ruzzene, M., A. M. Brunati, A. Donella-Deana, P. James, and L. A. Pinna. 1993. Regulation of p55^{c-fgr} tyrosine kinase by Csk and polycationic peptides. *Abstr. 18th Eur. Symp. Horm. Cell Regul.*
43. Sabe, H., B. Knudsen, O. Okada, S. Nada, H. Nakagawa, and H. Hanafusa. 1992. Molecular cloning and expression of chicken C-terminal Src kinase: lack of stable association with c-Src protein. *Proc. Natl. Acad. Sci. USA* **89**:2190-2192.
44. Skolnik, E. Y., C. H. Lee, A. Batzer, L. M. Vicentini, M. Zhou, R. Daly, M. J. J. Myers, J. M. Backer, A. Ullrich, M. F. White, and A. L. Et. 1993. The SH2/SH3 domain-containing protein GRB2 interacts with tyrosine-phosphorylated IRS1 and Shc: implications for insulin control of ras signalling. *EMBO J.* **12**:1929-1936.
- 44a. Songyang, Z., et al. Unpublished data.
45. Songyang, Z., S. E. Shoelson, M. Chaudhuri, G. Gish, T. Pawson, W. G. Haser, F. King, T. Roberts, S. Ratnofsky, R. J. Lechleider, B. G. Neel, R. B. Birge, J. E. Fajardo, M. M. Chou, H. Hanafusa, B. Schaffhausen, and L. C. Cantley. 1993. SH2 domains recognize specific phosphopeptide sequences. *Cell* **72**:767-778.
46. Taniguchi, T., T. Kobayashi, J. Kondo, K. Takahashi, H. Nakamura, J. Suzuki, K. Nagai, T. Yamada, S.-I. Nakamura, and H. Yamamura. 1991. Molecular cloning of a porcine gene Syk that encodes a 72-kDa protein-tyrosine kinase showing high susceptibility to proteolysis. *J. Biol. Chem.* **266**:15790-15796.
47. Tuveson, D. A., R. H. Carter, S. P. Soltoff, and D. T. Fearon. 1993. CD19 of B cells as a surrogate kinase insert region to bind phosphatidylinositol 3-kinase. *Science* **260**:986-989.
48. Valius, M., and A. Kazlauskas. 1993. Phospholipase C- γ 1 and phosphatidylinositol 3-kinase are the downstream mediators of the PDGF receptor's mitogenic signal. *Cell* **73**:321-334.
49. Waksman, G., S. E. Shoelson, N. Pant, D. Cowburn, and J. Kuriyan. 1993. Binding of a high affinity phosphotyrosyl peptide to the Src SH2 domain: crystal structures of the complexed and peptide-free forms. *Cell* **72**:779-790.
50. Whitman, M., C. P. Downes, M. Keeler, T. Keller, and L. Cantley. 1988. Type I phosphatidylinositol kinase makes a novel inositol phospholipid, phosphatidylinositol-3-phosphate. *Nature (London)* **332**:644-646.
51. Whitman, M., D. R. Kaplan, B. Schaffhausen, L. Cantley, and T. M. Roberts. 1985. Association of phosphatidylinositol kinase activity with polyoma middle-T competent for transformation. *Nature (London)* **315**:239-242.
52. Yi, T., and J. N. Ihle. 1993. Association of hematopoietic cell phosphatase with c-Kit after stimulation with c-Kit ligand. *Mol. Cell. Biol.* **13**:3350-3358.